



Digital Libraries

- What is a Digital Library?
- Research Issues [+ + + + +]
- The New NSF/ARPA/NASA Initiative

[\[Next Slide\]](#) [\[Talk abstract\]](#)
Hector Garcia-Molina

[\[Workshop Schedule\]](#) [\[Workshop Information\]](#) [\[Workshop Readings\]](#)



What is a Digital Library?

- Is it a computarized version of a current Library?
- Is it a system that gives access to all Digital Information?

[\[Next Slide\]](#) [\[Talk abstract\]](#) [\[Slides\]](#)
Hector Garcia-Molina

[\[Workshop Schedule\]](#) [\[Workshop Information\]](#) [\[Workshop Readings\]](#)



New research on systems for capturing data of all forms

- OCR page layout, segmentation and analysis software
- Speech recognition, audio segmentation and analysis software
- File conversion into editable, processable representations
- Broadcast capture & digitization (to generate multimedia data bases)
- Graphics understanding (image, drawing, graph recognition)
- Quality, fidelity maintenance

[\[Next Slide\]](#) [\[Talk abstract\]](#) [\[Slides\]](#)
Hector Garcia-Molina

[\[Workshop Schedule\]](#) [\[Workshop Information\]](#) [\[Workshop Readings\]](#)



New research on how to categorize and organize electronic information in a variety of formats

- Indexing, interpretation, classification and cataloging
- Multi-lingual indexing on content and citations
- Hypermedia structuring and linking of documents
- Graphical interfaces for knowledge representation
- Browsing technology for large knowledge spaces, news grazing

[\[Next Slide\]](#) [\[Talk abstract\]](#) [\[Slides\]](#)

Hector Garcia-Molina

[\[Workshop Schedule\]](#) [\[Workshop Information\]](#) [\[Workshop Readings\]](#)



New Research fundamental to the development of advanced software for searching, filtering, and summarizing large volumes of data, imagery, and all kinds of information

- Retrieval models for data, metadata, information, knowledge bases, evaluation methods
- Formal structures of documents and texts, query languages
- Intelligent text processing and document mgmt
- Feature-based image analysis and classification, pattern recognition
- Multi-pass retrieval algorithms (progressive refinement, feedback)
- Updatable indexing systems
- Spatial-temporal feature indexing of video
- Filtering, routing, alerting, selective dissemination of information
- Clustering, summarization, abstracting
- Natural language analysis for data extraction or representation
- Natural language generation systems
- Lexicon, thesaurus, concept space generation
- Adaptive/learning systems: connectionist, neural networks
- Context based pattern matching and retrieval of multimedia data
- Robust matching with noisy data, uncertainty, imprecision
- Fast search, query optimization



Research on visualization and other interactive technology for quickly browsing large volumes of imagery

- Pictorial feature recognition, image classification
- Human perception (visual, auditory)
- Multi-scale displays, zooming
- Data visualization (e.g., airflow, human genome)
- Interactive visualization control
- Use of simulation to improve visualization/description
- Navigation, hypermedia, retaining serendipity, guides/paths/tours
- Developing metaphors, usable virtual reality environment: suites of objects (peripherals, information types, relations, properties, views) that work together for each domain
- Sustaining rapid performance with regard to moving and manipulating large digital imagery data sets.

[\[Next Slide\]](#) [\[Talk abstract\]](#) [\[Slides\]](#)

Hector Garcia-Molina

[\[Workshop Schedule\]](#) [\[Workshop Information\]](#) [\[Workshop Readings\]](#)



Research on networking protocols and standards needed to insure the ability of the digital network to accommodate the high volume, bandwidth and switching requirements of a digital library.

- Network security
- Protocol design
- Data compression
- Ensuring scalability for large orders of magnitude increases in the number of simultaneous users

[\[Next Slide\]](#) [\[Talk abstract\]](#) [\[Slides\]](#)

Hector Garcia-Molina

[\[Workshop Schedule\]](#) [\[Workshop Information\]](#) [\[Workshop Readings\]](#)



New research leading to simplifying the utilization of networked databases distributed around the nation and around the world

- Knowbots/agents/mediators, intelligent gatekeepers
- Federated heterogeneous distributed object-oriented data and information base systems
- Personalized interactive news, magazine, and journal services
- Adaptable systems and services for disabled users, human augmentation
- Authentication, authorization
- Modeling, simulating usage, economics of access
- Collaboration technology with multimedia information interchange, multi-user editing, drawing, storage, retrieval, display, annotation, shared objects

[\[Next Slide\]](#) [\[Talk abstract\]](#) [\[Slides\]](#)
Hector Garcia-Molina

[\[Workshop Schedule\]](#) [\[Workshop Information\]](#) [\[Workshop Readings\]](#)



Research on individual and group behavioral, social and economic issues in digital libraries

- Intellectual property rights
- Privacy and security
- Impact of digital libraries on the conduct of science
- Publishing in a digital environment
- Charging mechanisms for copyrighted documents

[\[Next Slide\]](#) [\[Talk abstract\]](#) [\[Slides\]](#)
Hector Garcia-Molina

[\[Workshop Schedule\]](#) [\[Workshop Information\]](#) [\[Workshop Readings\]](#)



NSF/ARPA/NASA Initiative (**RUMORS**)

- Stanford
- Berkeley
- CMU
- Illinois
- Michigan
- UC Santa Barbara

[\[Talk abstract\]](#) [\[Slides\]](#)
Hector Garcia-Molina

[\[Workshop Schedule\]](#) [\[Workshop Information\]](#) [\[Workshop Readings\]](#)